It's good morning for you, some of you it's good evening. Welcome to our introduction of EAC-CPF version 2.0 brought to you by the Society American Archivists and the Technical Subcommittee on Encoded Archival Standards.

I'm the Co-Chair of TS-EAS. My usual work is being at Kommunalförbundet Sydarkivera in Sweden. With me today. I have the EAC-CPF team. Lead Marie Elia from the University of Buffalo, United States and team member Kerstin Arnold from Archives Portal Europe Foundation based in Germany even if Archives Portal Europe is based in all of Europe.

So welcome, and we'll start off right away. TS-EAS technical subcommittee on encoded archives standards. I think you get why we are using acronyms. We have on Youtube a short introduction to what we are and what we do.

So that is possible to look at. The presentation that we are using is already already uploaded to Github. So it's possible for you to use that one to get to click on the links. You need to download it because the links doesn't work in when you view on Github, but it's there for you to use.

At the same time we are available in a lot of different places. You will hear about some of them specifically today, but use the link in the presentation. When the work we do is following standards, we create the standards, and we are following a couple of revision schedules.

So everything we do. We have an annual rolling relation cycle from minor releases, that's part of the maintenance we do.

At the same time, being in the Society of American Archivists, our mother committee is the standards committee, and they have a guideline that every fifth year you should start big and make your revision of your standard.

So that is what we have just finished with EAC-CPF. At the same time, what we do is that in TS-EAS we have design principles. We have put up principles for how we design our standards and our XML schemas, and I'm just going to point you to the first one, simplicity comes first… everything should be simple to use and implement.

So that's our primary goal. The other nine principles, I will let you read them on your own, but you will see during today's presentation, that simplicity comes first has really gone through all the work we have done.

Getting everybody on the same page means that we decided that a short short introduction to what EAC-CPF is. So the acronym starts for encoded archival context, corporate bodies, person, and families and I'm using slides and content developed by a number of people who is also available to look at SAA's Youtube channel where you have a description and introduction to what all the EAS standards are. So that's also possible to look at.

But here is the short version of describing EAC-CPF. So the goal is to describe corporate bodies, person of families related to archival material, and with corporate bodies, that's actually everything that is not a person or a family.

So think about societies, small business, and everything that's a corporate body, and what we, what the schema is, is actually the technical representation of the International Standard of Archival Authority Records. ISAAR-CPF coming from this international council of Archivists.

So there you have all the elements, and when you read ISAAR-CPF, the names of and elements are the same that we are using.

We had to do some addition since we are talking about the technical format, but everything is there. We also run quickly, we'll end up in XML.

So XML, extensible markup language. What you will see throughout the presentation is that we are describing elements, the things in brackets, and attributes the ones with the @ sign, and that is the places where we put the information that we are encoding. These elements can be grouped in a hierarchical structure, meaning you have a parent element, and it has child elements. Child elements can have shared elements, and so on. But remember this structure.

To make everything smoothly. We have created an XML schema which tells how these elements that you are going to hear about today, and also the attributes can and have to be used.

And furthermore, we also give data types. So we define what the elemental attributes can contain. Is it text, is it numbers, and so on, so making sure that you can use them as they are expected to be used.

And the final thing is namespace, a tricky thing to actually really thoroughly explain. But it's the home space for elements and attributes in the schema.

So making sure that what we are doing has its own space, and all the things that we are describing is in this space. With that I'm actually going to jump directly into the XML so the structure of EAC is that you have a root element, the <eac>. And then you have two child elements.

You have control, which is an information about the XML file itself. You will hear a lot more about that later on, so I'm not going to really divel into that one.

The other part is CPF description. This is where you describe the entity. The corporate body, the person, or the family. Moving into the CPS description, we have three sub elements or three child elements.

Where the identity is the first one. It's required. It's there where you find information about the type of entity you are describing identifiers and you can have as many as you want. And also the name, and the name is also possible to give in all the ways that are needed. Something you will see later on.

Then we have description. So we want to describe these entities that we are creating a CPF record for. So you have two types of descriptive elements, some that are more purposed for being indexed; an address where you give existence functions, occupations.

You can give long biographical notes, and so on, and you can write as narrative as you want, or be as short as you want. So everything is possible.

This is a short example where you can see you can work both longer text and you can write, for example, placename is something that you might index so that's really short. And you have really short descriptions, or you can have a long description.

The last part is the relations. So when you are describing an entity, you really want to link it with other corporate bodies, persons, or families, you might want to relink it to the archival resource that this entity has created. Or you want to connect it with functions or something else. So that's what you do in the relations area.

When we look at this example of a relation, it's all centered about how the person we actually were describing is connected to other persons in the family. But I know. We will be touching in on this later on today.

So why do we have these relations? Well, we need to link the entity we have created with other things.

Actually we can link to other things being present on the Internet, because something we are connecting to might not be in our custody, so we actually point to something else.

And this, actually makes it easier to find the archival materials and get these big.

Some say the complex world of archival material makes it easier to understand. We can click ourselves around and find and make new connections and find new information.

So we have all the data we want to have. Then we need to create easy EAC-CPF documents, and there are some different ways of doing that.

The three usual examples are that you might have an XML editor to make this all by hand. You can have scripts that help you to create this, or if you actually have, this is part of your collection management system. So they actually create the EAC-CPF files for you.

When we have the files, we want to publish them, because we want other people and the researchers to actually see them. So then you can actually transform the XML into something that is browser readable, like HTML. Your collection management system can have a part which actually shows the patrons your descriptions or it can be a special system for this.

So giving some examples on where you can see EAC-CPF. In in being used is in 2012 you had a product name, Connecting the Dots: Samuel Johnson and His Circle at the Utton Library and Panicy Library, where they actually described Samuel Johnson and all his connections using EAC-CPF.

And they did amazing graphs of showing how people were connected and everything. We have SNAC the social networks and archival context which is present in the U.S.

Where you also had other people or other countries have supplied information. There you also can see people and how they are connected; and they are using EAC-CPF as the format for describing the people or all the corporate bodies or families.

The last one is actually Archives Portal Europe, where all the archive creators have an EAC-CPF record showing you the information about the creators.

With that quick start and loop into EAC-CPF, I'm going to hand it over to Marie and Kirsten, so I know Maria starting…please take it off

And just before I do, in case anyone came in a little bit later, I'm gonna put the link to the PDF of the slides one more time in the chat. As Karin said, you'll have to download the PDF to click on the links, then you'll be able to see this.

So the focus of today is on the updates and how everything works. So I'm going to be very brief to summarize the revision process.

So just for context, we're at just over 10 years since the adoption of the original version of EAC-CPF, which was in 2010.

The first major revision which the team began working on in 2017 they released. Say, they because I haven't joined yet. Least the minor update the following year, and then place the call for comments on that version.

So the resulting overhaul of the standard was submitted to the Standards Committee and the SAA Council earlier this year, and approved in August.

The overall goals for the revision were to simplify where possible, align with EAD where it makes sense, implement features and solutions based on user feedback and clearing up any unused components.

And you can find a very detailed explanation of the revision process on the EAC-CPF Website on the left. On the sidebar menu there's EAC-CPF 2.0 background that if you go to that page and click on revision notes you'll see a very long detailed explanation of the process.

But I'm gonna give an overview of some of these changes.

The goal of EAD alignment is to meet the user community's needs and make it easier to maintain and use the EAC-CPF schema.

So as part of this alignment with EAD, some elements were renamed, and some were renamed to be more precise.

So ideally these changes make the schema easier to use and understand

Also as part of the alignment, some elements were removed and replaced sometimes by a new equivalent term, and sometimes by transformation. So it would be a combination, maybe an element, and attribute now to represent the same information

To simplify with regards to elements. EAC-CPF 2.0 bundles elements in these 2 ways. Elements of the same type can be bundled with a wrapper as plural elements and elements with different concepts can be grouped in element sets. and we'll see examples of these as we go

Further regarding the simplification, there's a prescribed order for elements within the parent elements, so prioritizing required a non repeatable elements

Restrictions were relaxed for some element, content and attribute values. So these are just some examples for the element agency code.

The ISO constraint is relaxed to text, and for the country, language and script code attributes the ISO standards constraint is relaxed to name token.

I will mention a couple more things, but we'll be covering this again in more detail.

Starting with control. But these are some examples. And we'll have some examples of encoding.

But these are some examples of how the new attributes align with EAD. And these option elements, as well.

The removal and replacement of some attributes regarding external namespaces. As Karin was mentioning earlier about namespace. This also furthest the alignment with EAD.

And finally, there are these five new or replacement global attributes. We'll have a detailed explanation of internal and external referencing later, again, with examples. So I will just move on to Kerstin to start with control.

Okay. Thank you very much. So Marie gave you the quick overview. We're now gonna dive a little bit in more detail into specific sections of EAC-CPF, respectively specific types of information that you might want to encode.

We're gonna start with control. And Marie already mentioned that we have transformed some elements into attributes specifically in the control section where we essentially kind of have all the information that we need to manage the EAC-CPF file in itself.

We felt like this would help us to kind of get the process more streamlined, and there are 4 elements that we have treated that way, so we've got the maintenance status and the publication status of the is telling you something about the the processing stage so to say of the EAC-CPF file.

Then we have the maintenance event type, and the agent type. They're both in the maintenance history of the EAC-CPF file, and all 4 of them currently come with a list of limited values, and you can see with the asterisk that we are already kind of reviewing this again because next to the EAC-CPF revision we also have started a major revision of EAD, so the wheels are turning already, even though we only just published EAC-CPF 2.0.

We also have a few new attributes and possibilities to internally reference from one part of EAC-CPF to the other, and in turn that results into some sub elements of the control section being emphasized more,

So ideally you would kind of use the control section for example, to provide information about any rules or conventions that you have been following by creating the EAC-CPF instance, any sources that you have used, and then also kind of detailing who was involved in the creation, or later on the updates of the EAC-CPF file.

So all of these elements now have an ID attribute, so that they can be pointed to, and we have reference attributes that you will find all over the place in their descriptive part of EAC-CPF in order to point to those parts in control.

Just as kind of a use case. I'm moving away a little bit for a second from the XML encoding.

So you could say, I want to start a new EAC-CPF record from a working group and document one when I did it.

So you would want to have a unique ID for the record.

You will want to mention who has created it, and you might also want to record a date when it was created, and if you go to the next slide, we already see how this would look like in the EAC-CPF encoding and that is essentially everything that you need to have.

So all the mandatory information that you need to have when creating an EAC-CPF file in the control section, so you can see it's not that much.

There are lots of other possibilities of what you can include but kind of to start with, you would have a kind of relatively short control section with the main information.

We can spin this a little bit further. So if you wanted to say that the document is in draft, but it's already public, then we can use the audience attribute, saying that everything that is in here is external and can be used by anyone, but we can also indicate the publication status as being in process so that users already know that there might be more information added in a later stage.

Marie already mentioned those attributes, as well. I might also want to point to specific data standards that I'm using in the context of the record.

So if I'm using specific date encoding standard like the ISO Standard 8601 throughout the record when, for example, talking about birthdays or death dates of the person that I'm describing then I can point to that in the control section which also helps in validation.

But also helps others to understand how your data has been encoded.

This how this would look like in control. So you can see there's just one element, the control element, and then all the extra information is essentially kind of packed away into attributes. All of these attributes are optional, so you don't have to use them, but you can use them.

Looking a little bit into the maintenance event as one example, as that is also part of the mandatory information that we would want to have. So you might want to make an internal note in the record about what you did, why and when, and you can do all of that in the maintenance event element.

So you have the name. You have a date and maybe a time when something has happened, and you can also include a description of what you have done, and if we go to the next slide you can see how this looks like in the EAC-CPF encoding, so we've got the maintenance event as the parent element in this context, and you can see that we have 3 child elements.

The agent which gives you the personal. In this context. The group that has created the EAC-CPF file got event date time, which gives you the date and more specifically at the time.

If you needed in the normalized version, and then you have event description where you can essentially include any longer text that you might want to provide.

Going into the section of linking and referencing, and we can just move on to the next slide.

We have essentially kind of 2 different ways of referencing, and we are looking at external referencing in terms of pointing out of the EAC-CPF to a resource that is somewhere else and that different ways to do this what we have included in this new version is a set of 3 attributes to point to external vocabularies or ontologies.

So something like Wiki data, something like via something like the Library of Congress subject headings.

All of that could be included in those 3 attributes. We have referencing to sources that you have used in creating the EAC-CPF file and the source element can for this have a sub element called reference, which you can use for either analog sources or including the @ href attribute also point to online sources.

And the same principle with reference is also available throughout the EAC-CPF file in the descriptive part. If you want to kind of provide any external resources for more context.

These are some encodings of how this could look like. So at the top we have a resource where we point specifically to catalog entry at the Barack Obama Presidential Library.

The second example is using the vocabulary referencing so you can see that in this case, we are pointing to Wiki data and the entry about Barack Obama on them, and then in the relation part.

The third example, essentially kind of gives some more context information by pointing to a Wikipedia page.

Next to that, we also have the internal referencing, and also mentioned the id attribute which is available in every element that there is an EAC-CPF apart from the root element EAC itself and going along with this is the target attribute, so you essentially kind of have something where you can point to with ID, and something where you can point from with the target attribute.

So you can essentially kind of reference from any element within an EAC-CPF document to any other element within the same document, and then we have these more specific target attributes, so to say, that I briefly mentioned earlier in the context of control we go to the next slide we'll see what these look like.

Actually this is just an example still for target and ID. So, just having a look at this at the top, you can see that in the context of occupation, a place is just generally named, and then the targets is pointing to an actual place element where you can then have more detailed information about the place, like giving the plays a role and including address information.

Now we come to the more specific target attributes and we have generally called that assertion description. It is something that is coming from the context of SNAC.

So this picking up on community feedback, and if you are familiar with SNAC, you might know that essentially there are different people, editors, who have run through a certain training working on the same EAC-CPF file, so that's so in that context it's really essential to know who did what when, and based on which sources and this is what we are doing here.

So we have created 3 new attributes, maintenance, event, reference that allows you to point to who did what, when source reference that allows you to point to a source, and convention declaration reference that allows you to point to any rule that has been applied by creating the content.

That's especially useful in case of conflicting statements. So you might already have had a description of a person, and then, a few years later, someone does a little bit more research, and, for example, new date of birth appears, and you might want to include that in your EAC-CPF description but you might also want to identify that this is something that is new and might be something that still needs to be confirmed.

And this is where these attributes come in.

And these are again a few examples for this, and you can see that we can use the ID attribute, because that's something that we always need to have in order to point to other elements.

And then the relation at the bottom essentially kind of includes the source reference as an example and the maintenance event reference ss a second example, where you can see that this is pointing to the first 2 entries.

Similarly, we also have a concept that is called local types, and that's something that already wasn't in EAC-CPF. In the previous version. But we are again kind of emphasizing that concept a little bit more so in some elements in the descriptive part you might have a local type where you can include any typing that you might use in your context.

But as we also think about EAC-CPF is something that you might want to share.

For example, in context of an aggregator like SNAC or Archives Portal Europe. It's of course, good. If you can tell people what you actually mean by these local types so similarly, to what we have with the convention declaration reference and the other reference attributes.

We also have local type declaration reference option. So essentially in the control section, you have local type declaration, and you can point to, for example, or locally used vocabulary or terminology list that you might want to use throughout the EAC-CPF description, and then in the actual descriptive part for example, with the name entry you can point back to that local type declaration,

Now we go into the section of looking at specific pieces of information that you might want to encode, and how this is done in EAC-CPF 2.0 and we start with the names as being one of the most important aspects in describing a person, a corporate body or a family.

We have renamed the the grouping element. So Marie mentioned earlier that we have plural elements and set elements in order to group different child elements together in the context of name entries.

The initial name was name Entry Parallel, which throughout the revision process we found out that this is very specific to a certain way to encode names in the U.S. based context but doesn't necessarily translate into other international contexts, so we decided to rename that into name entry set and we also transformed a few of the sub elements in the context of encoding names, and again, we essentially kind of transformed elements into attributes.

And if we go to the next slide we have an encoding example. How this would look like now, so for example in the terms of names, you might want to reference a specific rule or convention you've been following, for example if you've always had the last name first, and a comma and the first name.

And In the name entry itself. You can, for example, declare that a specific name entry is the preferred form.

So if someone went by a certain name that isn't necessarily their birth name, you can use the preferred form, attribute to indicate that, and you can also indicate whether that name entry is an authorized name or is a name that is an alternative for specifically in the context of corporate bodies.

If you have an institution, like this one here, which is an international one you might have the name in different languages, and all of them might be authorized.

But if you have an institution that is placed only for example, in Spain, then you might probably have the the Spanish name as the author as one and a translation of the name into English, might just be an alternative form of that.

Just going into the name entry set so showing a few variations. What you can do with this or in a name, and a name entry set you can have several main entries, and, for example, you can put name entries in different languages in there.

So each language would have their own name entry tag. And you can kind of include specific information with regard to what type of language you have the name in.

You might also want to indicate the script, if that is relevant in the context, and you can again kind of play with the preferred form, true or false, and with authorized or non authorized names respectively, also using the local type, if you wanted to to indicate that something is the native language of the person, and something is a translation.

The next piece of information that we want to look at is place encoding.

This is something where there was a mix between things that we included as something new, because we thought for that it is essentially kind of just moving on to the more digital way of doing things, and things that we have aligned with EAD.

So if you go to the next slide, we are essentially just differentiating between 2 different use cases where we talk about places.

We have one part of the description where you can encode places in full, so we have a plural element places, and then we have other bits and pieces in the description where we might just want to name a place as as one thing and then you can kind of reference to the more detailed description if available, and if wanted.

Place in itself in the full description requires at least one of the following sub elements.

So either the name of the place, which is highly recommended, the role of the place you can define, so you can define a place as being the birthplace of a person or a workplace of a person. We have a possibility to encode a physical address, and then one thing that we added new is the possibility to next to that also encode digital address information.

For example, website on email and we adopted the possibility to encourage geographic coordinates from the EAD.

Furthermore, you can also add date, informational description to place element if that is relevant in the context of the overall description.

This is just an example on how this could look like if you were to use all the sub elements of the place element.

So we start with the place name. You can see that it also is a country code that you can indicate. Specifically, if you're having a place that might appear in different countries, you can give the place role geographic coordinates.

The address is broken down into different address lines, and each address line can be typed so you can specify which part of the address is the street and maybe the house number which is the name of the city, which is supposed to code in this case and the same principle also applies to the contact information where we have contact line elements. And in this case we have included the homepage.

The next one is date encoding, and we have specifically looked at the possibility of encoding uncertain, approximate, and ongoing dates, which was something that you couldn't do in the previous version of EAC.

So this is again something where we picked up on user feedback. So we now have a possibility to include a statement of uncertainty. So we adopted the certainty attribute from EAD.

We have a possibility to include normalized uncertainty. So we have kind of extended the use of ISO 8601 to the current version with the extended date, time, format, and we also have a possibility to indicate a dates status with the new attribute status and you can use the values unknown or ongoing in this context with specifically in context of describing persons or institutions, could be valuable.

This is how this could look like, so you've got the certainty attribute which is really just an open, text attribute. You can include anything that you want in there. Then a date range where we have the start age unknown, and for the end date we have an approximate date, and this is also expressed in the normalized version, and then in the third example, we have a date set, and you can see that the date range that is included in there is ongoing.

So this is something that is still very current. And the last one is relations. So Karin mentioned relations are at a very prominent place in the context of EAC-CPF, and we adapted a little bit how this can be encoded.

So if we go to the next slide, please, we can have a look at that.

So we decided to not differentiate with the element between what type of entity we are relating to. So we only have the general relations element now. And within relation you can name target entity, so name the other person that you're pointing to, or name the resource that you're pointing to, because that has been created by the person that you are describing, and we have have the attribute target type going with that where you can indicate what type of entity it is that you are pointing to, and this is picking up essentially on what we had previously.

So you can half the target type, person, family, corporate body, or more generally agent. You can have a target type function and target type resource.

Furthermore, we have included 2 new elements where you can more specifically describe the relation that the 2 entities have with each other.

Relation type is the more general one. So that gives you kind of a general relationship description between the entity that you are describing in the EAC-CPF and the other entity that you're pointing to.

And target role gives you the possibility to more specifically describe what role the targeted entity plays towards the entity that you are describing, and on the next slide we have an iconic encoding example for how this could look like.

So here we are, pointing to Paul Arendt in the context of describing Hannah Arendt.

You can see that there also are a few of these vocabulary attributes in there again.

So you are pointing to kind of a standardized description of Paul Arendt, and with the relation type we say that there is a family relation between both of them and in the target role.

We can more specifically say that Paul Arendt is apparent, or we could also say, is the father of Hannah Arendt.

This is where I hand it back to Marie for the documentation part.

And before I start, I'm actually gonna put these links in the chat and as well they are linked from the presentation. But just for easy reference.

So this is a brief overview of the highlights and changes that we think will be most noticeable and useful to the user community. But you can see everything in detail on the EAC-CPF Website. Again details of the revision but the link to the Tag library, the link to the TS-EAS. Github site where you will also find the best practice guide

So our EAC-CPF Team member, Iris has created a short tutorial video which is linked there in the SAA Youtube Channel, which shows you how to navigate through the best practice guide.

So we have shown, peppered throughout examples. Here we have examples on Github. There's a fully encoded example of an EAC-CPF record, an extended example. There's a brief example, and there are 2 just sort of sample records that have more than generic information in them.

But we very much want to see how you were using EAC-CPF; we would like the best practice guide to include your use cases, examples, and you can submit examples and communicate with us issues through the Github site or by contacting us.

We want again, the best practice guide in our documentation to be useful and the community effort.

So my big congratulations to the EAC-CPF team for doing this work. And really we were really proud, having it being approved, and being announced on the third of August this year.nSo it's available, and it's ready for everybody to use.

So as Marie said, please put your questions in the chat, and I would like to try to do this magic of stopping the recording so many thanks for attending.

Please feel free to ask questions.